Tasking on Natural Statistics of Infrared Images

Todd Richard Goodall, Alan Conrad Bovik, *Fellow, IEEE*, and Nicholas G. Paulter Jr., *Fellow, IEEE*

***Abstract*— Natural scene statistics (NSSs) provide powerful, perceptually relevant tools that have been successfully used for image quality analysis of visible light images. Since NSS capture statistical regularities that arise from the physical world, they are relevant to long wave infrared (LWIR) images, which differ from visible light images mainly by the wavelengths captured at the imaging sensors. We show that NSS models of bandpass LWIR images are similar to those of visible light images, but with different parameterizations. Using this difference, we exploit the power of NSS to successfully distinguish between LWIR images and visible light images. In addition, we study distortions unique to LWIR and find directional models useful for detecting the halo effect, simple bandpass models useful for detecting hotspots, and combinations of these models useful for measuring the degree of non-uniformity present in many LWIR images. For local distortion identification and measurement, we also describe a method for generating distortion maps using NSS features. To facilitate our evaluation, we analyze the NSS of LWIR images under pristine and distorted conditions, using four databases, each captured with a different IR camera. Predicting human performance for assessing distortion and quality in LWIR images is critical for task efficacy. We find that NSS features improve human targeting task performance prediction. Furthermore, we conducted a human study on the perceptual quality of noise- and blur-distorted LWIR images and create a new blind image quality predictor for IR images.**

***Index Terms*— NSS, LWIR, "Halo effect," hotspot, NU, TTP.**

I. Introduction

LONG WAVELENGTH Infrared (LWIR) images have many uses in industry, military, medicine, and science. Non-destructive testing uses thermal imagers for detecting defect locations in manufactured materials, thereby allowing for better quality control [1]. Unmanned Airborne Vehicles (UAV) and security cameras often couple a thermal imager with a visible light (VL) camera to enhance night vision for scouting and to improve automatic threat detection over large distances [2]. Firefighters carry handheld imagers while scouting for critical burn points in burning buildings and possible thermal hazards [3]–[5]. Thermographers use high-resolution thermal imagers for detecting inflammation, irregular blood-flow, and tumors [6].

Manuscript received November 22, 2014; revised August 2, 2015; accepted October 17, 2015. Date of publication October 30, 2015; date of current version November 20, 2015. This work was supported by the National Institute of Standards and Technology under Grant 70NANB12H283. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Damon M. Chandler.

T. R. Goodall and A. C. Bovik are with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: tgoodall@utexas.edu; bovik@ece.utexas.edu).

N. G. Paulter, Jr., is with the Materials Measurement Science Division, National Institute of Standards and Technology, Gaithersburg, MD 20899 USA (e-mail: paulter@nist.gov).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2015.2496289

A broad theme of this paper is the development and practical application of Natural Scene Statistic (NSS) models of LWIR images. NSS models describe statistical regularities that are observed on images taken of the natural world.[1] Examples of NSS of visible light images include the $\frac{1}{f}$ behavior of the amplitude spectrum [7], [8], the sparse coding characteristic of visual cortical filters in response to natural image stimuli [9], and the Gaussianity exhibited by visual signals following band-pass filter and adaptive gain control operations [10]. Early cortical processing in higher mammalian visual systems appears to have adapted to these natural statistics [7], and much research into biological visual functioning has been guided by the "efficient coding" hypothesis, which assumes that visual neurons have adapted to efficiently encode natural visual stimuli [11], [12].

Given their widespread use and application, LWIR images have been well studied. Mooney characterized sources of spatial noise [13] and the effect of noise on minimum resolvable temperature differences (MTD) as a function of frequency [14]. Lopez-Alonso further characterized spatial noise in IR images by using Principle Components Analysis (PCA) to separate spatial and temporal noise from a sequence of frames [15]. This led Pezoa and Medina to model the non-uniformity (NU) noise common in LWIR images expressed in the frequency domain [16] as distinct from independent spatial noise. Using this NU model, Pérez *et al.* measured and compared the efficacy of several non-uniformity correction (NUC) algorithms [17] and developed methods for extracting the structure of the underlying fixed-pattern noise [18].

Although NSS have proven to be highly successful tools in applications on visible light images, the development and use of similar models has not been nearly as widespread on LWIR images. Morris *et al.* compared LWIR image statistics with natural visible light image statistics, and found that the spectral power of LWIR images is more "heavy-tailed" and that LWIR wavelet histograms are generally peakier, likely due to the characteristic spatial smoothness of infrared images. Kaser [19] and Goodall and Bovik [20] modeled the fit of the BRISQUE [21] and NIQE [22] image quality models to LWIR images showing that these visible light models fit reasonably well to LWIR image data. However, as we show later, the statistics of visible-light and LWIR are predictably different. To measure NU, noise, blur, and changes in brightness, Amon *et al.* developed four Image Quality Indicators (IQIs) [4], [5]. To measure NU in LWIR images, Hayat *et al.* introduced a Roughness Index computed

[1] In essence, captured photographically, of any real-world scenes, including both man-made and naturally occurring objects.

(a) (b) (c)

Fig. 1. Example artifacts of distorted LWIR images. (a) Non-uniformity. (b) "Halo effect". (c) Hotspots.

using discrete spatial derivatives [23]. An improved index, the Effective Roughness Index [24], measures this roughness index using a high-pass image. To our knowledge, we have described the extent of previous work in modeling or utilizing NSS on LWIR images.

LWIR images are certainly 'natural' in the sense that we use the term, and understanding and modeling the NSS of LWIR images has the potential to drive the development of new applications. These applications include identifying distortions present in any given LWIR image, enhancing images to reduce the degree of distortion, and for producing more accurate and reproducible comparisons of the performance of thermal imagers. Since the final receiver of LWIR images is often the human observer, human tasks can be improved by incorporating NSS that capture the response of the human visual system. Important distortions that we study here include non-uniformity (NU) noise, ferro-electric "Halo effects," sensor noise, JPEG artifacts, blurring, and hotspots.

A. Characteristics of LWIR Imagers and Images

Thermal imagers have a spectral sensitivity in the 7 μm to 14 μm wavelength band. Thermal imagers are sensitive to the radiation emitted by objects in the scene and the background radiation reflected by those objects. The relationship between the irradiation collected at the sensor and the temperatures of imaged materials is nontrivial. For example, the material and surface properties of an object alter its emissivity. Similarly, the reflective properties of an object will vary the amount of background radiation reflected by the object and subsequently collected by the imager. All this variability can lead to errors in the measurement of an object's temperature.

Although both cooled and uncooled thermal imagers exist, uncooled imagers are the most widely adopted and are the ones considered herein. Also known as Focal Plane Array (FPA) imagers, they are designed to use either an array of resistors (called a microbolometer) or an array of ferro-electric ceramic transistors. The microbolometer works by measuring changes in resistance corresponding to changes in temperature. The ferro-electric technology operates by measuring a temperature differential across a pyro-electric material, which is refreshed by a constantly spinning wheel, called the "chopper." As a result of the "chopper," the images obtained by these detectors exhibit additional or more severe artifacts such as the "Halo effect," which is lowering their adoption in the market. Unlike cooled detectors, the typical design of uncooled imagers does not allow adjustment of their sensor integration time, thus they usually capture images at a lower frame rate than cooled imagers. Overall, the main advantage of uncooled imagers over their cooled counterparts is their lower power consumption, cheaper components, and size [25], [26].

Non-uniformity (NU) noise, as exemplified in the image in Fig. 1(a), is a distortion specific to LWIR images. NU is an additive fixed-pattern noise, which appears as a grid-like or striping pattern. These patterns result from manufacturing defects, dark current, and segmented sensor capture areas [15], [16], [27].

The "Halo effect," depicted in Fig. 1(b), is another distortion, which occurs mostly in thermal cameras equipped with ferro-electric sensors. This effect causes the region surrounding a bright object to grow darker and it causes the region around dark objects to grow lighter [25], [28]. This effect can be caused by both the physical operation of cameras containing ferro-electric sensors and by the back-reflection of IR illumination sources. The "chopper" that modulates the signal for ferro-electric detectors fails to entirely shield incoming infrared light, which leads to overcompensation when subtracting the differential response from the average signal, thereby producing a halo. Reflective materials situated next to highly emissive materials within the scene have also been shown to produce a similar effect [29].

LWIR images commonly contain hot-spot areas exhibiting only small variations in image intensity, that arise from highly emissive objects that stand out from the background as in Fig. 1(c). In general, LWIR images contain many smooth surfaces as a result of temperature diffusion. Hot-spots are less a distortion than a symptom of the environment, but they still produce interesting statistical regularities worthy of study.

Other unique degradations of LWIR images not covered herein include radiometric distortions, geometric distortions, noise from reflective materials, and the history effect. Radiometric distortion refers to non-linear mapping of thermal energy to pixel values in an image, which may destroy relevant sensitivity information. Geometric distortions occur when the sensors in the FPA are mis-aligned, causing blur. As discussed previously, materials imaged with an infrared camera are often assumed to be only emissive, but they can also be reflective causing false inferences. Lastly, heat in the LWIR band can fluctuate faster than frame rate, which can be difficult to detect given the physical limits of infrared sensors. Geometric distortions are specific to individual imagers, radiometric distortions appear during the capture process, reflective noise

(a) (b) (c) (d)

Fig. 2. Example images from 4 LWIR databases. (a) NIST. (b) KASER. (c) MORRIS. (d) OSU.

TABLE I
CAMERAS ASSOCIATED WITH THE 4 LWIR DATABASES

Database	Detector Material	Lens Diameter (mm)	Bit Depth	Resolution	Sensitivity (mK)
OSU	Ferro-Electric	75	8	360x240	100
MORRIS	Microbolometer	50	8	384x288	60
NIST	Microbolometer	10	14	640x480	55
KASER	unknown	unknown	8	640x480	unknown

measurements require knowledge of the captured objects, and the history effect is a time-varying distortion. These distortions and accompanying side information are not currently available for study thus they are not included here.

The same distortions that occur in visible light images can of course also occur in LWIR images. For example, blur may arise from camera destabilization, especially in handheld devices, non-optimal lens focus moving the depth of field away from the object of interest, or object motion. Sensor noise may be induced by light sensitivity based on the integration times of the sensors. Over and under-exposure can occur as a function of exposure time, or from quantization or local saturation. JPEG distortion such as blocking artifacts and blurring can also be present, since most thermal cameras utilize the JPEG compression format.

B. LWIR Image Sources

Our study of the NSS of LWIR images has benefited from the availability of four separate IR image databases that we denote as NIST [30], KASER [19], MORRIS [31], and OSU [32]. Example images from each database are provided in Fig. 2. The NIST database includes 180 images from indoor office and home environments, where each image contains multiple hot objects. The KASER database includes 37 images from outdoor environments, each captured using unknown camera models and suffering various distortions including non-uniformity (NU), blur, and noise. The MORRIS database contains a total of 288 indoor and outdoor images of urban environments including cars, pedestrians, and buildings. Finally, the OSU database contains a total of 262 images captured by a surveillance camera monitoring pathway intersections on the Ohio State University campus. Gathering a diverse set of images from a diversity of cameras allows for better generalization and characterization of the NSS of LWIR images.

The general characteristics of the uncooled thermal cameras associated with each database are listed in Table I. The KASER database contains pristine images and images contaminated with mixtures of unknown distortions, making them less suitable than NIST or MORRIS for natural scene analysis. The images obtained from the camera type in OSU required processing by an additional non-linearity (using a log transform) in order that the NSS followed the same regularities as images captured from the microbolometers. The capture method used by the OSU imager obtains pixel values that follow a non-linear function of luminance, like power. Since this non-linear function is unknown, we use OSU separately from our NSS modeling analysis.

All images were linearly mapped to the range 0 to 1 for comparability and ease of applying artificial distortions consistently. This does not change the image statistics beyond normalizing them to this range.

C. Distortion Models

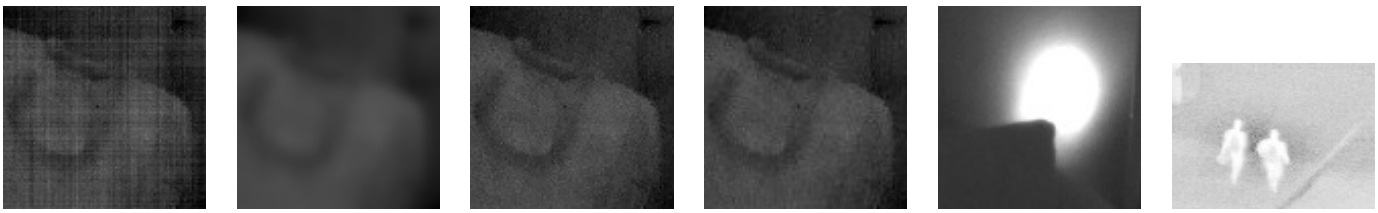
We next describe the generative noise models used to create distorted LWIR images. Pezoa and Medina developed a model of non-uniformity that can be used to artificially distort pristine images [16]. Based on a spectral analysis of NU, they proposed the model

$$|\tilde{I}(u,v)| = B_u \exp\left(\frac{-(u-u_0)^2}{2\sigma_u^2}\right) + B_v \exp\left(\frac{-(v-v_0)^2}{2\sigma_v^2}\right)$$

$$\angle \tilde{I}(u,v) \sim \mathrm{U}[-\pi,\pi]$$

where $\tilde{I}$ is the Fourier Transform representation of the noise image, $B_u = B_v = 5.2$, $\sigma_u = \sigma_v = 2.5$, and where $\mathrm{U}[a,b]$ denotes the uniform distribution on $[a,b]$. The severity of NU can be controlled by scaling the dynamic range using a standard deviation parameter σ_{NU}.

Three levels of distortion each from JPEG, Additive White Noise (AWN), non-uniformity, and blur were applied to the NIST and MORRIS databases. These two databases were



(a) $\sigma_{\mathrm{NU}} = 0.025$ (b) $\sigma_{\mathrm{b}} = 3.0$ (c) $\sigma_{\mathrm{AWN}} = 0.025$ (d) JPEG (quality 80%) (e) Hotspot patch (f) "Halo effect" patch

Fig. 3. Images in 3(a), 3(b), 3(c), and 3(d) depict the third (most severe) distortion level for NU, blur, AWN, and JPEG. Figures 3(e) and 3(f) depict patches selected for hotspots and halos.

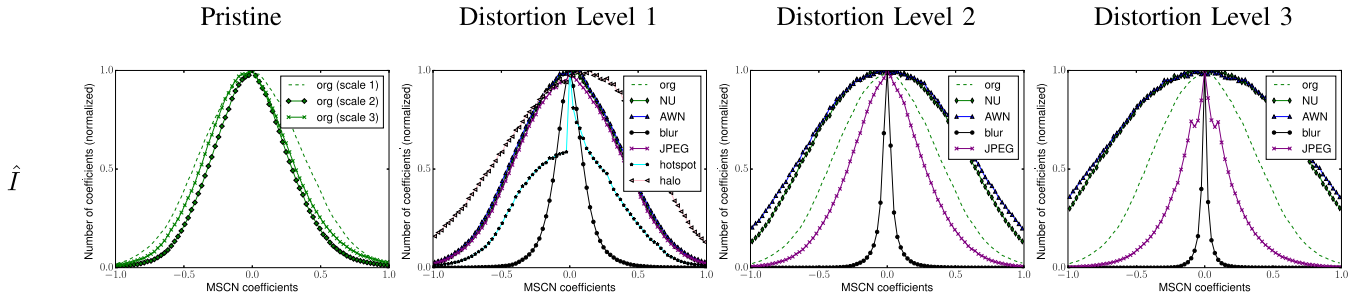


Fig. 4. MSCN histograms of 468 ROIs, one ROI extracted from each image in the NIST and MORRIS LWIR image databases. The ROIs were 64 pixels by 64 pixels and nominally centered in the image. The left-most figure depicts the scale-invariant behavior of the MSCN statistics of non-distorted images. The remaining three figures depict increasing distortion levels at the first scale. The terms org, NU, AWN, blur, JPEG, hotspot, and halo refer to pristine images, images with NU distortion, images with AWN distortion, images with blur distortion, images compressed with JPEG, hotspot image patches, and halo image patches, respectively.

chosen since the images were the least visually distorted. JPEG images were generated at the 100, 90, and 80 percent quality settings corresponding to distortion levels 1, 2, and 3 producing average respective bit rates of 3.6, 1.0, and 0.5 bpp. Levels 1, 2, and 3 of additive Gaussian white noise and non-uniformity were generated using $\sigma_{\mathrm{AWN}} = \{0.0025, 0.01375, 0.025\}$ and $\sigma_{\mathrm{NU}} = \{0.0025, 0.01375, 0.025\}$ (recall the gray-scale range is $[0, 1]$). Levels 1, 2, and 3 of blur were generated using a Gaussian blur kernel with scale parameters $\sigma_{\mathrm{b}} = \{1, 2, 3\}$. The third (most severe) distortion levels for JPEG, AWN, NU, and blur are depicted in Fig. 3.

Two additional distortions, hotspots and the "Halo effect," were cropped by hand. The "Halo effect" occurs naturally in the images in the OSU database. Davis's method [33], which is based on background subtraction and morphological techniques, was used to isolate moving objects (often people) in the images. Since not all objects extracted using this method exhibited the "Halo effect," patches with a clear visible "Halo effect" were isolated by hand. A total of 188 example patches were thus selected from the OSU database for use here. Hotspots were isolated by hand from the NIST and MORRIS databases. A total of 135 hotspot patches including people, environmental hazards, and other miscellaneous objects were extracted. Since these two distortions were not were not generated, these distortions only provide 1 distortion level. Example patches for hotspot and "Halo effect" distortions are shown in Fig. 3.

II. NSS of LWIR Images

A. Processing Model

In a pioneering deep study of the statistics of visible light images, Ruderman observed that applying a local bandpass filter combined with a non-linear operation to a natural image has a decorrelating and Gaussianizing effect [10]. Given that highly successful Image Quality Assessment (IQA) models have used this property to measure distortions in visible light images, we analyze this Gaussianizing property further for LWIR images. Given an input luminance image, I, its Mean-Subtracted Contrast Normalized (MSCN) coefficients are defined by

$$\hat{I}(i, j) = \frac{I(i, j) - \mu(i, j)}{\sigma(i, j) + C}$$

over spatial indices with $i \in 1, 2 \ldots M$, $j \in 1, 2 \ldots N$, where M and N are the image height and width, respectively, C is a constant that prevents instabilities when the denominator tends toward zero. The factors μ and σ are weighted estimates of the local luminance mean and standard deviation given by

$$\mu(i, j) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l} I_{k,l}(i, j)$$

and

$$\sigma(i, j) = \sqrt{\sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l} (I_{k,l}(i, j) - \mu(i, j))^2}$$

where $w = \{w_{k,l} | k = -K, \cdots, K, l = -L, \cdots, L\}$ is a 2D circularly-symmetric weighting function sampled out to 3 standard deviations and normalized to unit volume.

Histograms of these MSCN coefficients for natural LWIR images appear similar (Gaussian) to those of visible light images, as shown in Fig. 4. To compute these histograms with minimal distortions, coefficients were pooled by selecting one centrally located region of interest (ROI) of size 64×64 from

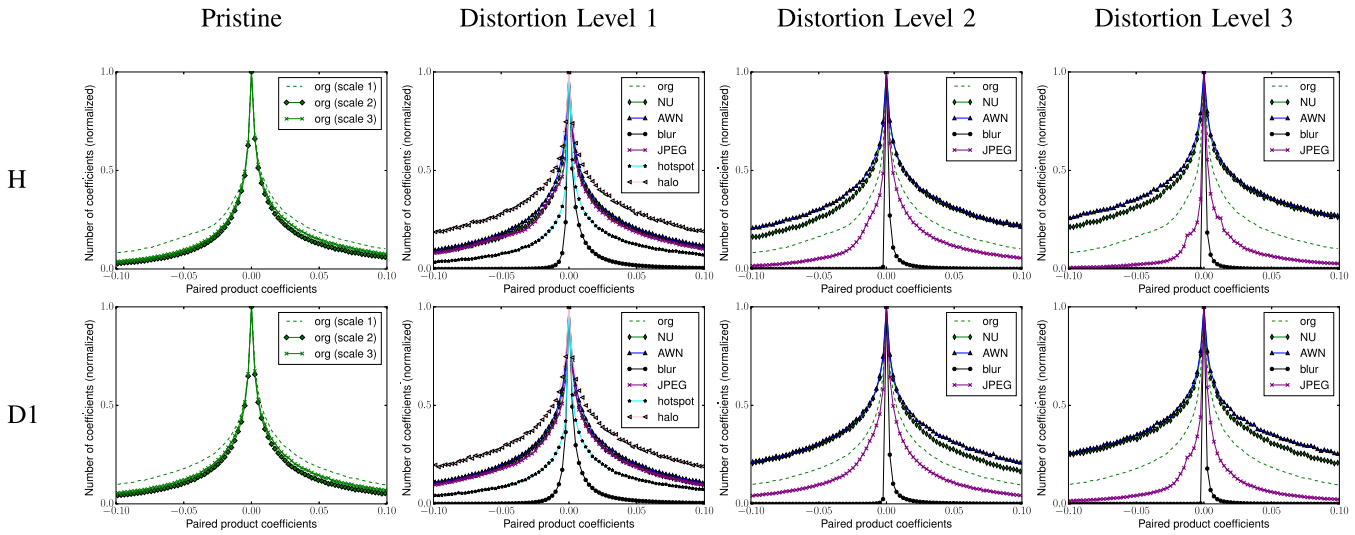


Fig. 5. Paired product histograms of 468 ROIs, one ROI extracted from each of the images in the NIST and MORRIS databases. The ROIs were 64 pixels by 64 pixels and nominally centered in the image. The left-most column depicts the scale-invariant behavior, and the remaining 3 columns depict increasing distortion levels at the first scale. See [34] for the complete comparison.

468 images taken from the NIST and MORRIS databases. These histograms demonstrate that undistorted LWIR images are statistically regular.

The three distortion levels for JPEG, AWN, NU, and blur as produced in Sec. I-C undergo this same MSCN transformation and center ROI selection for fair comparison in Fig. 4. The hand-cropped hotspot and halo distortions are included as Level 1 distortions. Non-uniformity and additive white Gaussian noise perturb the regularity observed in pristine ROIs, producing wider histograms, since they unnaturally increase the overall variation of images. By contrast, JPEG and blur create an artificial high degree of image correlation, producing thinner histograms. The hotspot artifacts produce asymmetry in the histogram, while the "Halo effect" histogram is wider, having an apparent skew.

In the Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE) model [21], neighboring MSCN coefficients are multiplied to produce paired product coefficients. Four directional coefficient products are computed at each coordinate (i, j)

$$\begin{aligned}
\mathrm{H}(i,j) &= \hat{I}(i,j)\hat{I}(i,j+1)\\
\mathrm{V}(i,j) &= \hat{I}(i,j)\hat{I}(i+1,j)\\
\mathrm{D1}(i,j) &= \hat{I}(i,j)\hat{I}(i+1,j+1)\\
\mathrm{D2}(i,j) &= \hat{I}(i,j)\hat{I}(i+1,j-1).
\end{aligned}$$

The paired products are included for analyzing the directional behavior of both the statistical regularity and perturbations thereof for LWIR images.

Histograms of these paired product coefficients were produced following the same procedure used for the MSCN images. Fig. 5 depicts the H and D1 paired product histograms. These distortion-free histograms are similar across three scales, as also observed for visible light images. NU and AWN distortions produce relatively wider histograms, while blur and JPEG produce thinner histograms, similar to the trends observed in the MSCN histograms. Hotspots produce thinner histograms and halo artifacts produce wider histograms. A more complete comparison of these histograms can be found in [34].

In an interesting extension of BRISQUE called the Derivative Statistics-based QUality Evaluator (DESIQUE) model [35], the MSCN histograms are supplemented by seven log-derivative coefficients computed by differencing the logarithms of the magnitudes of neighboring MSCN coefficients. The goal of these coefficients is to provide higher sensitivity to high frequency noise. The following function is defined

$$J(i,j) = \log\left(\left|\hat{I}(i,j)\right| + K\right)$$

where K is a stabilizing constant, and the log-derivative coefficients are computed as

$$\begin{aligned}
\mathrm{PD1}(i,j) &= J(i,j+1) - J(i,j)\\
\mathrm{PD2}(i,j) &= J(i+1,j) - J(i,j)\\
\mathrm{PD3}(i,j) &= J(i+1,j+1) - J(i,j)\\
\mathrm{PD4}(i,j) &= J(i+1,j-1) - J(i,j)\\
\mathrm{PD5}(i,j) &= J(i-1,j) - J(i+1,j)\\
&\quad - J(i,j-1) - J(i,j+1)\\
\mathrm{PD6}(i,j) &= J(i,j) + J(i+1,j+1)\\
&\quad - J(i,j+1) - J(i+1,j)\\
\mathrm{PD7}(i,j) &= J(i-1,j-1) + J(i+1,j+1)\\
&\quad - J(i-1,j+1) - J(i+1,j-1)
\end{aligned}$$

Following the described ROI extraction procedure for both pristine and distorted images, histograms of these Log-Derivative coefficients are plotted in Fig. 6. These histograms seem to indicate less differentiation among NU, AWN, and halo distortions, but good sensitivity to JPEG and blur distortions, when compared to MSCN and paired product histograms. Again, see [34] for a more complete comparison of these log-derivative histograms.

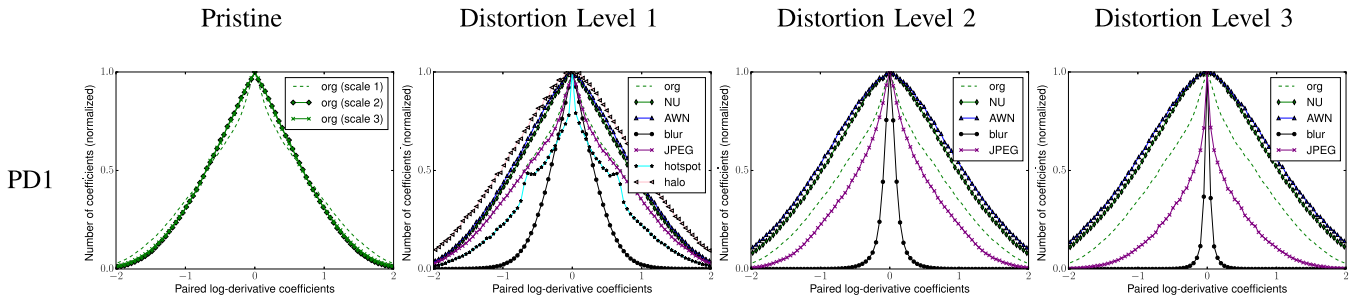


Fig. 6. Paired Log-Derivative histograms of 468 ROIs, one ROI extracted from each of the images in the NIST and MORRIS databases. The ROIs were 64 pixels by 64 pixels and nominally centered in the image. The left-most column depicts the scale-invariant behavior, and the remaining 3 columns depict increasing distortion levels at the first scale. See [34] for the complete comparison.

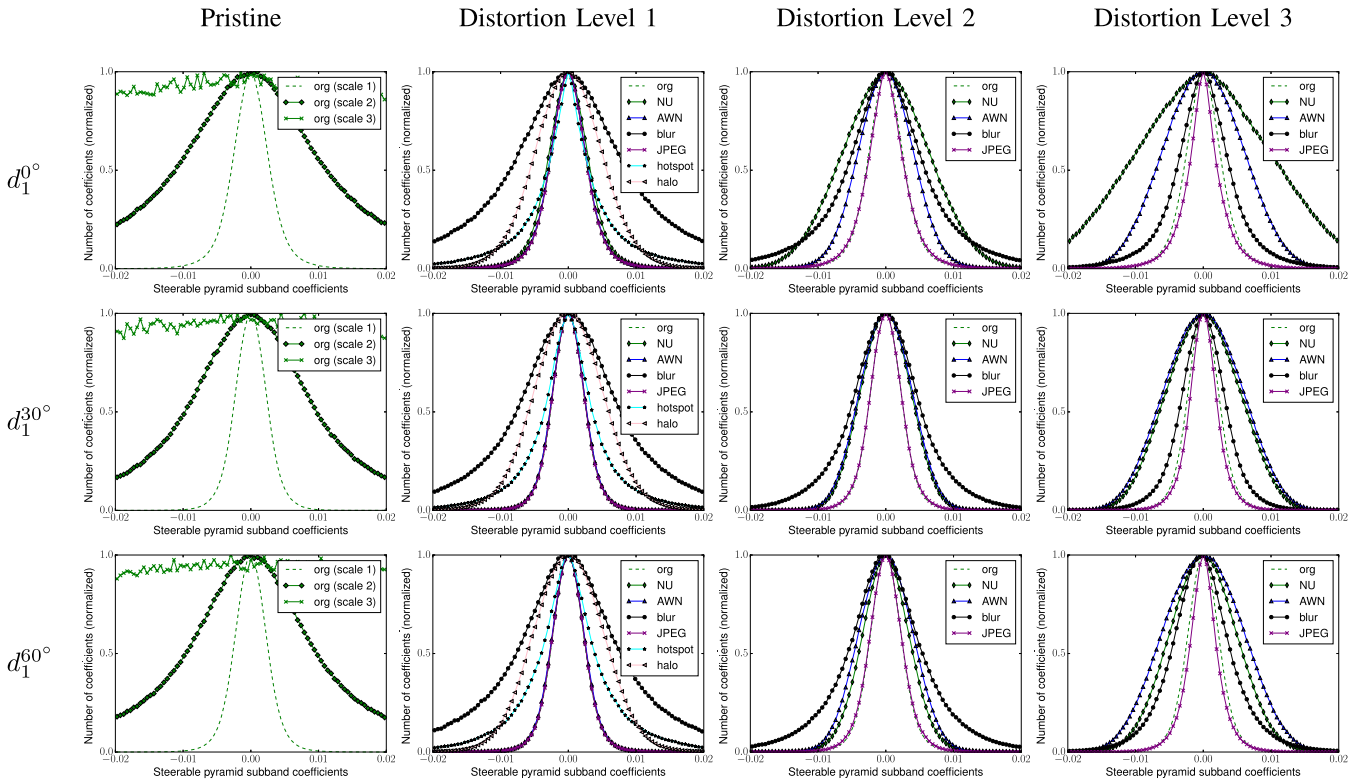


Fig. 7. Steerable Pyramid histograms of 468 ROIs, one ROI extracted from each of the images in the NIST and MORRIS databases. The ROIs were 64 pixels by 64 pixels and nominally centered in the image. Only 3 orientations are shown, but all provide useful and distinct information. The left-most column depicts differences in standard deviation across scale. The remaining three columns indicate a qualitative change in histogram shape for changes in distortion level. See [34] for the complete comparison.

Since the MSCN, paired product, and paired log-derivative coefficients operate on normalized high-pass images, they do not directly, explicitly describe band-pass characteristics. A perceptual model that captures these band-pass characteristics is the divisively normalized steerable pyramid decomposition, which mimics perceptual neurons in the early stages of the human visual pathway, capturing information over multiple orientations and scales [36]–[38].

Using the same pooled ROI extraction procedure, histograms produced from normalized steerable pyramid coefficients are plotted in Fig. 7. Each band is denoted d_α^θ where α denotes the scale and $\theta \in \{0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ\}$. The pristine coefficients depict different histogram widths for each scale, corresponding to the different relative energies for high, mid, and low frequency bands. Analyzing distortion behavior, NU has a distinctively large standard deviation in the horizontal and vertical subbands, d_1^0 and d_1^{90}, which follows from the striping behavior of NU. All other distortions appear to be invariant to orientation, but distinct when compared to the natural image histograms. See [34] for a full comparison of the divisively normalized steerable pyramid histograms.

B. Feature Models

A parametric General Gaussian Distribution (GGD) [39] has been used to model the MSCN, Paired Log-Derivative, and steerable pyramid subband coefficients. The associated

TABLE II
FEATURE SUMMARY FOR MSCN (f), PAIRWISE PRODUCTS (pp), PAIRED LOG-DERIVATIVES (pd), AND STEERABLE PYRAMID SUBBANDS (sp) FOR THE FIRST SCALE

Feature ID	Feature Description	Computation Procedure
f_1 - f_2	Shape and Variance	GGD fit to MSCN coefficients
f_3 - f_4	Shape and Variance difference	GGD fit to right and left halves of MSCN coefficients
pp_1 - pp_4	Shape, mean, left variance, right variance	AGGD fit to H pairwise products
pp_5 - pp_8	Shape, mean, left variance, right variance	AGGD fit to V pairwise products
pp_9 - pp_{12}	Shape, mean, left variance, right variance	AGGD fit to D1 pairwise products
pp_{13} - pp_{16}	Shape, mean, left variance, right variance	AGGD fit to D2 pairwise products
pd_1 - pd_2	Shape and Variance	GGD fit to PD1 pairwise log-derivative
pd_3 - pd_4	Shape and Variance	GGD fit to PD2 pairwise log-derivative
pd_5 - pd_6	Shape and Variance	GGD fit to PD3 pairwise log-derivative
pd_7 - pd_8	Shape and Variance	GGD fit to PD4 pairwise log-derivative
pd_9 - pd_{10}	Shape and Variance	GGD fit to PD5 pairwise log-derivative
pd_{11} - pd_{12}	Shape and Variance	GGD fit to PD6 pairwise log-derivative
pd_{13} - pd_{14}	Shape and Variance	GGD fit to PD6 pairwise log-derivative
sp_1 - sp_2	Shape and Variance	GGD fit to $d_1^{0^\circ}$ subband
sp_3 - sp_4	Shape and Variance	GGD fit to $d_1^{30^\circ}$ subband
sp_5 - sp_6	Shape and Variance	GGD fit to $d_1^{60^\circ}$ subband
sp_7 - sp_8	Shape and Variance	GGD fit to $d_1^{90^\circ}$ subband
sp_9 - sp_{10}	Shape and Variance	GGD fit to $d_1^{120^\circ}$ subband
sp_{11} - sp_{12}	Shape and Variance	GGD fit to $d_1^{150^\circ}$ subband

GGD probability density function is

$$f(x;\alpha,\sigma^2) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} exp\left(-\left(\frac{|x|}{\beta}\right)^{\alpha}\right)$$

where

$$\Gamma(x) = \int_0^{\infty} s^{x-1}e^{-s}ds.$$

An Asymmetric Gaussian Distribution (AGGD) [40] has been used to effectively model to the paired product coefficients. The pdf is

$$f(x;\upsilon,\sigma_l^2,\sigma_r^2) = \begin{cases} \frac{\upsilon}{(\beta_l+\beta_r)\,\Gamma(\frac{1}{\upsilon})} exp\left(-\left(\frac{-x}{\beta_l}\right)^{\upsilon}\right) & x < 0 \\ \frac{\upsilon}{(\beta_l+\beta_r)\,\Gamma(\frac{1}{\upsilon})} exp\left(-\left(\frac{x}{\beta_r}\right)^{\upsilon}\right) & x \geq 0 \end{cases}$$

where β_l and β_r are given by

$$\beta_l = \sigma_l \sqrt{\frac{\Gamma\left(\frac{1}{\upsilon}\right)}{\Gamma\left(\frac{3}{\upsilon}\right)}}$$

and

$$\beta_r = \sigma_r \sqrt{\frac{\Gamma\left(\frac{1}{\upsilon}\right)}{\Gamma\left(\frac{3}{\upsilon}\right)}}$$

respectively.

The parameters (α, σ^2) of the GGD model fit can be estimated using the technique described in [39]. The parameters (υ, σ_l^2, σ_r^2) of the AGGD model fits can be estimated using the moment matching technique described in [40]. Another parameter, η, given by

$$\eta = (\beta_r - \beta_l)\frac{\Gamma\left(\frac{2}{\upsilon}\right)}{\Gamma\left(\frac{1}{\upsilon}\right)}$$

is computed for each product image using the estimates of the other parameters. Therefore, the best-fit model of each set of paired product coefficients yields 4 features (η, υ, σ_l^2, σ_r^2).

Since the hotspot images exhibit asymmetric histograms, negative and positive MSCN coefficients were measured separately. Negative and positive coefficients correspond to the left and right halves of the histograms. Therefore, four parameters (α_l, σ_l^2, α_r, and σ_r^2) were extracted from the MSCN coefficients. The differences in value between the left and right halves, $\alpha_r - \alpha_l$ and $\sigma_r - \sigma_l$, are used to capture the asymmetry. A overview of the MSCN (f), paired product (pp), paired log-derivative (pd), and steerable pyramid subband (sp) features is provided in Table II.

To visualize the clustering of the features over three scales, the features for each distortion class were projected into a 2D space using Principle Component Analysis (PCA) as depicted in Fig. 8. The distorted images appear to cluster in this projection, reasonably preserving their class groupings.

A boxplot comparing the features in Table II between pristine LWIR images and pristine visible light images is provided in Fig. 9. A total of 29 pristine visible light images were obtained from the LIVE Image Quality Assessment Database [41]–[43]. The MSCN shape parameter, f_1, is not significantly different between visible and LWIR images when using 95 percent confidence intervals. Comparing f_3, we can infer that LWIR images provide more symmetrically shaped MSCN histograms with 95 percent confidence.

The mean parameter, η, for each of the paired product features differs between LWIR and visible light images. Additionally most of the standard deviation parameters, σ_l and σ_r, differ between the modalities. Most shape parameters for paired products do not appear to differ between LWIR and visible light images. By contrast, most of the shapes and standard deviation parameters for pd and sp are significantly different from visible light images. Individual parameter differences are

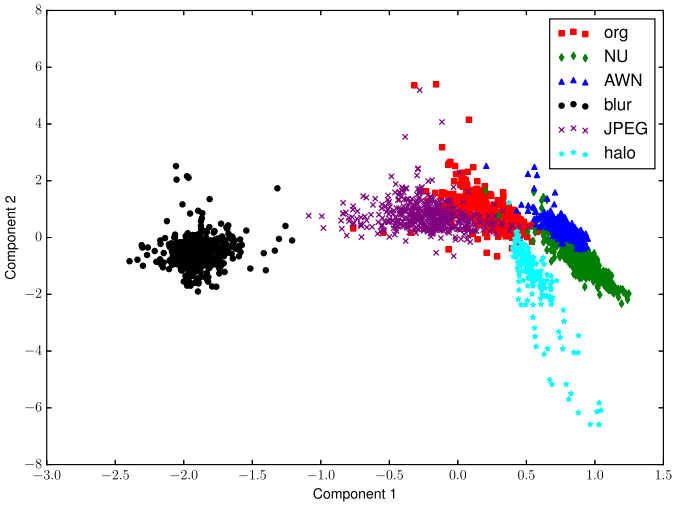


Fig. 8. A total of 46 features over 3 scales yields 138 features per image, projected here into 2D space using PCA. Despite the total explained variance ratio being 0.734, distorted images cluster away from the natural images. Hotspots were not included due to how far they projected. Here, org refers to pristine images.

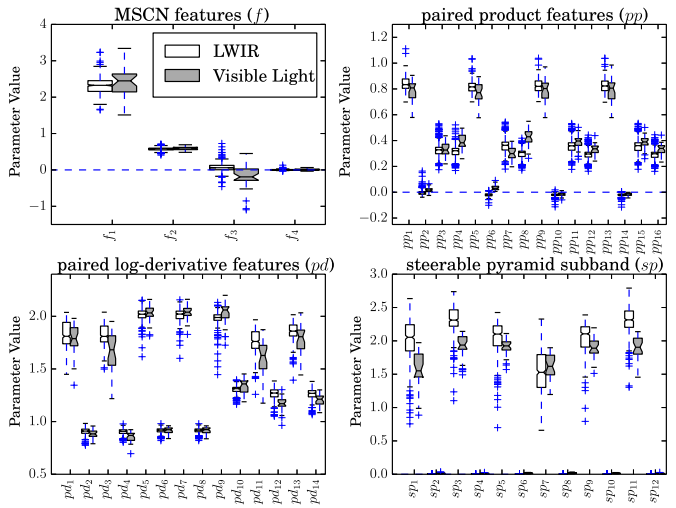


Fig. 9. Box plot comparison of features between natural LWIR and natural visible light images. The notches indicate 95 percent confidence intervals about the median. These plots show that while the generic NSS models for visible light and LWIR images are similar, their specific parameterized models differ.

expected to exist given the number of features being compared; however, a difference certainly exists between the two groups, as further explored in Sec. III-A. Note that these differences may be due to either differences between imager technologies or the differences between the physics of visible light and IR fluctuations in a scene.

C. NIST Descriptors

Previous work by the National Institute for Standards and Technology (NIST) has produced four Image Quality Indicators (IQIs) [3]–[5], which are described as brightness (B), contrast (C), spatial resolution (SR), and non-uniformity ($\widehat{\text{NU}}$) defined as follows:

- **B** is the average of the luminosity intensities:

$$\text{B} = \frac{1}{MN}\sum_{i\in N}\sum_{j\in M} I(i,j)$$

- **C** is defined as RMS contrast:

$$\text{C} = \sqrt{\frac{1}{MN}\sum_{i\in N}\sum_{j\in M}(I(i,j)-B)^2}$$

- **SR** (cycles/pixel) is computed by

$$\text{SR} = \int_0^{f_c} (MTF_{curve}(u) - \text{NEM})\, du$$

where $MTF_{curve}(u)$ is the modulation transfer function defined by the Butterworth filter

$$\widetilde{H}(u) = 1/\left(1+\left(\frac{u}{W_n}\right)^4\right)$$

of order 2. The cutoff frequency is

$$f_c = W_n\left[(1-\text{NEM})/\text{NEM}\right]^{0.25}$$

where NEM = 0.02861 is the Noise Equivalent Modulation.

- $\widehat{\text{NU}}$ is given by $\widehat{\text{NU}} = \mu/\sigma = \text{B/C}$, the SNR of the image.

As currently defined, the SR statistic, which depends directly on the parameter W_n, is not implementable. This dependency on W_n assumes that any loss of spatial resolution can be modeled based on the response of a Butterworth filter. According to Morris *et al.* [31], the log of the radial averaged spectral power of LWIR images can be well described as following a GGD probability law. Unfortunately, this fit does not generalize when distortions are present in an image, thus a 10th order polynomial approximation was used to yield a much better fit to the radial magnitude of the frequency spectrum. Overall, the IQIs provide a total of 13 features that are extracted from each image. Unlike the other features, the IQI features are not model based, but rather are sample statistics.

III. Tasking on NSS

In this section, we study the practical usefulness of the LWIR NSS and IQI features just described for solving six different visual LWIR tasks. First, in Sec. III-A, we use the NSS to discriminate between visible light and IR images. Second, in Sec. III-B, we use these features to develop a measure of NU on LWIR images. Third, in Sec. III-C, we devise a method to determine presence of the "Halo effect." In the fourth task in Sec. III-D, we utilize NSS to automatically predict LWIR Targeting Task Performance (TTP) of human experts. Fifth, in Sec. III-E, we describe a human study that obtained subjective quality scores on LWIR images, and show that the NSS features are highly predictive of subjective image quality. Lastly, in Sec. III-F, we show how NSS features can be used to create localized distortion maps that can aid the identification of local distortions such as hotspots and occurrences of the "Halo effect."

TABLE III
ACCURACY OF THE VISIBLE-IR DISCERNER FOR 1000 TRAIN/TEST ITERATIONS

Model	Input image type	Accuracy
Mahalanobis	Visible light	1.000
	Infrared	0.915
Random Forest	Visible light	0.994
	Infrared	0.970

A. Discriminating IR From Visible Light Images

The Visible-IR Image Discerner (VIID) can be used to effectively distinguish IR from visible light images using only NSS. For this test, only pristine visible light and pristine infrared images are used to provide a direct discrimination between modalities. As shown in NIQE, the NSS covariance structure converges for pristine images using only a few images. Previous analysis using NIQE on IR images [20] shows a similar convergence for IR images. The degree to which these NSS covariance structures differ between IR and visible light images is unclear, thus a test was devised to find whether natural IR images can be discriminated from a corpus of both IR and visible light images.

A total of 108 pristine visible light images were obtained from the reference images contained in the CSIQ [44], LIVE [41], DRIQ [45], VCL@FER [46] databases. A corresponding 108 IR images were obtained from MORRIS and NIST by random selection. A total of 1000 randomized train/test sets were evaluated, with each set containing 80 visible light training images, 80 infrared training images, 28 visible light test images, and 28 IR test images. For each set, a classifier was trained only on the 160 training images and used to label the IR and visible light test images. Thus, hit rates for visible light and IR images were recorded for each set. We use a random forest classifier [47] and a classifier based on the Mahalanobis distance measure [48].

The Mahalanobis distance, given by

$$D(x;\mu,\Sigma)=\sqrt{(x-\mu)^T\,\Sigma^{-1}\,(x-\mu)}$$

is computed using $D(x;\mu_{\mathrm{VL}},\Sigma_{\mathrm{VL}})$ on visible light images and $D(x;\mu_{\mathrm{IR}},\Sigma_{\mathrm{IR}})$ on infrared images. The means and covariance matrices μ_{VL}, Σ_{VL} and μ_{IR}, Σ_{IR} were computed using the visible light and IR training images respectively. A Mahalanobis classifier is developed to discriminate between VL and IR using

$$\mathrm{L}=\arg\min\left\{D(x;\mu_{\mathrm{VL}},\Sigma_{\mathrm{VL}}),D(x;\mu_{\mathrm{IR}},\Sigma_{\mathrm{IR}})\right\}$$

such that $\mathrm{L}=0$ and $\mathrm{L}=1$ correspond to visible light and infrared images respectively.

The mean hit rates (accuracy) from the 1000 train/test sets are reported in Table III. The results of the Mahalanobis-based classifier show that the natural IR and natural VL feature spaces are indeed distinct. In addition, the random forest classifier was used to produce a top-20 feature importance map as depicted in Fig. 10. The paired product group pp has a higher importance than group f, given that f does not appear in the top 20 ranked feature list. The vertical paired product at the finest scale (scale 1) is significantly more important than the features corresponding to the other directions, H, D1, and D2 at scales 2 and 3. Using the transposes of the visible light images as input in the above tests confirms that this phenomenon is intrinsic to the visible light images, i.e., infrared images appear to exhibit fewer prominent horizontal edges, lines or structures than visible light images, which is related to the overall smoothness observed in infrared images.

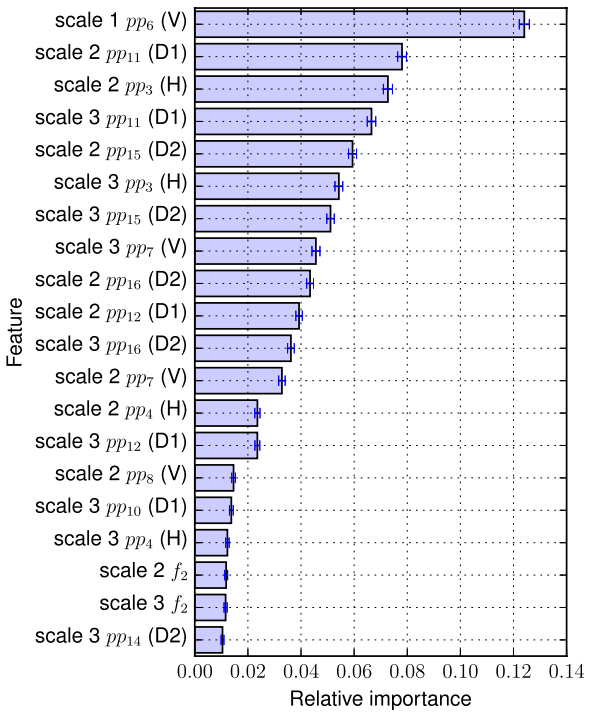


Fig. 10. Top 20 important features used for the Visible-IR discernation task. The importances were averaged over 1000 train/test iterations, shown here with 95% confidence intervals.

B. Measuring NU

In NUC algorithms, producing a no-reference estimate of the NU in an image is essential [17]. Such an estimate of NU can be used as both an input into a correction algorithm and as a performance measure for that correction algorithm. State-of-the-art methods for estimating the magnitude of NU include the Roughness index, Effective Roughness Index, and SNR. LWIR images commonly contain both fixed pattern noise and additive white noise, and the level of both types of noise should be reliably estimated.

The most common method for estimating NU is the spatial SNR of the image defined as μ/σ where σ and μ are the standard deviation and mean pixel intensities within a user-defined area. Another common and popular method, the Roughness [23] index, is

$$\mathrm{Ro}(I)=\frac{\|h_1*I\|_1+\|h_2*I\|_1}{\|I\|_1}$$

where h_1 is the 1D differencing filter with impulse response $[1,-1]$, $h_2={h_1}^T$, and $\|\cdot\|_1$ is the L_1 norm. The Effective Roughness [24] index is

$$\mathrm{ERo}(I)=\mathrm{Ro}(g*I)$$

where g is a high-pass filter, with the additional modification that the L_2 norm is used in place of the L_1 norm.

Two weaknesses of current NU estimation approaches are their inability to capture spatial structure and their reliance on the assumption of grid-like patterns of fixed-pattern noise, which, in reality, can instead be striping [17]. Additionally, these approaches generally assume that NU is the only distortion within the image. Often, other noise is present that can seriously hinder the effectiveness of estimating NU.

A new approach that we have devised to measure NU utilizes the proposed NSS features listed in Table II. To compare against existing NU estimation techniques, we degraded the images in the MORRIS and NIST databases each with non-unifomity and additive white noise. The standard deviations of the non-uniformity and additive white noise are produced by sampling $\sigma_{\text{NU}} \in \text{U}[0.025, 0.0025]$ and $\sigma_{\text{AWN}} \in \text{U}[0.025, 0.0025]$ respectively. Three categories of degradations were produced, those with just NU distortions (set 1), those with AWN distortions (set 2), and those with combined NU and AWN distortions (set 3). The two types of noise are thus independently embedded into the source image for each of the three sets.

Using these three sets of degraded images, we compared the predictive performance of each NU metric. A Support Vector Regressor (SVR) was used to learn features for estimating σ_{NU} in sets 1 and 3 and for estimating σ_{AWN} in sets 2 and 3. Each set was tested and trained independent of the other sets. Each set was split into non-overlapping subsets: 80 percent for training, and 20 percent for testing. The Spearman Rank Correlation Coefficient (SRCC) was used as a measure of non-linear monotonicity between the actual and predicted values, and (Pearson's) linear correlation coefficient (LCC) was used as a measure of linear correlation between actual and predicted values. Random 80/20 splits were produced and scored 1000 times, and the median SRCCs and LCCs are fully tabulated in [34].

The results show that the NSS feature groups, f, pp, pd, and sp, produced better predictors of AWN and NU as compared to Ro, ERo, and the IQIs for each of the three sets. Combinations among these NSS feature groups do not increase predictor accuracy significantly. The f group has lower correlation when compared to the other feature groups that contain a sense of directionality. It is important to note that the IQIs have almost no correlation with σ_{NU}, especially for set 3. Additionally, these IQIs were a comparatively mediocre predictor of σ_{AWN} on sets 2 and 3.

Individual features can be found that correlate highly with horizontal striping NU and vertical striping NU separately. The sp group features show significant correlation with directionality, with vertical striping effects being highly correlated with the d_1^0 subband standard deviation, and horizontal striping effects being highly correlated with the d_1^{90} subband standard deviation. The paired product features indicate a similar oriented correlation, the horizontal paired product σ_r, pp_4, correlates highly with vertical striping, and the vertical paired product σ_r, pp_8, correlates highly with horizontal striping. This feature-level correlation with the magnitude σ_{NU} of striping is useful not only for detecting the presence of a striping artifact, but also for measuring how much striping is present.

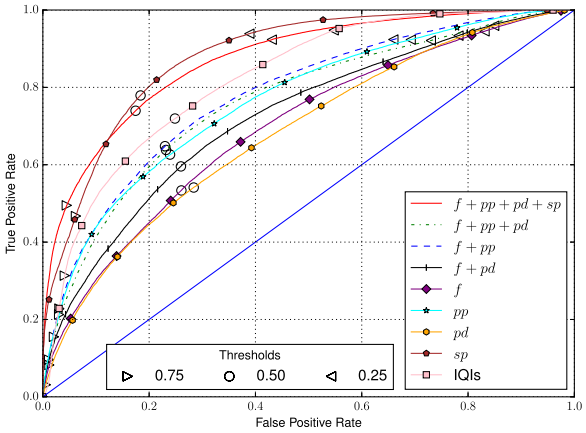


Fig. 11. ROC indicating the ability of NR algorithms to sort patches as either containing halos or as non-halo patches. Curves computed from 1000 train/test iterations using 415 total patches from the OSU dataset without content overlap. The 0.75, 0.5, and 0.25 thresholds are indicated for reference.

C. Discriminating the "Halo Effect"

The authors of [28] developed a person detector that used the statistical gradients of estimated halos to enhance the detection task. To our knowledge, no methods exist for detecting halo artifacts in LWIR images.

To study how well the "Halo effect" can be discriminated using our feature models, two sets of image patches (with and without halos) were constructed using a mixture of background subtraction and manual selection, as described in Sec. I-C, followed by manual classification. Most of the image patches were of size 110×110. A total of 415 image patches were contained in both sets, with 227 image patches being halo-free, and 188 patches containing halos.

AWN and NU distortions were applied to each patch in both sets to reduce the dependence on the correlation between "Halo effect" and the level of other common noise distortions. Each of these 415 image patches thus contained two artificial distortions in addition to the halo effect distortions. The distortion magnitudes $\sigma_{\text{NU}}, \sigma_{\text{AWN}} \in \text{U}[0.0025, 0.025]$ were randomly sampled and used as the variance of the white noise and non-uniformity distortions for each patch. The intervals for this uniform distribution were selected to scale the distortion from a just-noticeable to a significant difference.

Given these two distorted sets, those containing halos and those without, we devised a binary classification task. We split the dataset into two non-overlapping subsets: 80 percent for training and 20 percent for testing. A Support Vector Classifier (SVC) was used to map the features between two classes. Random 80/20 splits were produced and classified with associated class probability estimates 1000 times. The ground truth and classifier probabilities per feature group are appended each iteration with replacement.

Receiver Operating Characteristic (ROC) curves for the binary classification task using the proposed feature groups and the IQIs are shown in Fig. 11. These curves

TABLE IV
AREAS UNDER THE ROC CURVES IN FIG. 11

NR Feature Set	Area Under ROC Curve
$f + pp + pd + sp$	0.795
$f + pp + pd$	0.711
$f + pp$	0.723
$f + pd$	0.675
f	0.651
pp	0.699
pd	0.639
sp	0.795
IQIs	0.735

are computed on the aforementioned appended ground truth and probability pairs. The areas under the ROC curves are reported in Table IV. The proposed sp feature group and combinations of sp with the NSS feature groups achieved highest performance. The IQIs achieved low performance for this discrimination task. Speculatively, both directionality and band-specific tuning are important when detecting the "Halo effect," which follows from the general asymmetry and inhibitory-type behavior of this artifact.

D. TTP of Firefighters and Hazards

Researchers at NIST conducted a study involving firefighters whose task was two-fold [4]. First, given an LWIR image, the expert determined whether a hazard was present. Second, if a hazard was present, the expert was asked to identify the location of the hazard. This study was broken up into two phases. The phase 1 study used 4500 images. These images were created by degrading 180 pristine images. Five different levels of degradation corresponding to each IQI were generated and 25 sets of the four IQIs were used (for a total of 100 unique arrangements of the five values of each of the four IQIs). These 25 sets were deemed sufficient to represent the defined IQI space (5^4). Phase 2 used 55 sets of the four IQIs (for a total of 9900 images). The larger number of sets served also to extend the range of IQIs to include more extreme values. Note that the IQIs in this study were used as distortion-generating settings, allowing for direct measurement of distortion with TTP.

In this study, the experts were given a stimulus image, and tasked to either identify the location of the environmental hazard by clicking on it, or by indicating that there is no distortion. To better isolate detectability, we converted the dataset into patches centered about the hazards. Images with no hazards were discarded. Next, only the scores of observers that attempted to identify the location of the present environmental hazard were kept. Hits and misses were measured depending on whether the cursor click was near the hazard. The probability of hit was computed over all observers. By modifying the dataset in this way, SRCC and LCC correlations between target quality and target detectability could be more directly measured.

Using the probability of hit, the NSS quality features, and the IQIs, we used an SVR to estimate TTP. As a way of comparing the features, the median SRCC and LCC coefficients are reported in Table V over 1000 iterations.

TABLE V
MEDIAN SRCC AND LCC BETWEEN ACTUAL AND PREDICTED TTP FROM 1000 ITERATIONS

NR Feature Set	SRCC	LCC
$f + pp + pd + sp$	0.665	0.671
$f + pp + pd$	0.640	0.646
$f + pp$	0.582	0.601
$f + pd$	0.609	0.613
f	0.504	0.527
pp	0.562	0.582
pd	0.566	0.568
sp	0.340	0.367
IQIs	0.621	0.630

Fig. 12. Example stimulus.

Combinations of features provide the best estimators of TTP, with the combination of all natural features providing the highest correlations for TTP. Note that the IQIs in Table V use the 13 features, while the degradations to the images provided in the study made modifications based on the original 4 parameters.

E. Blind Image Quality Assessment of LWIR Images

We conducted a lengthy and sizeable human study, the results of which we used to assess how well NSS-based blind image quality prediction models designed for LWIR images correlate with subjective quality scores. A collection of 28 indoor and outdoor images were selected from the NIST and KASER databases as "pristine" images. Artificial blur and additive white gaussian noise distortions were applied to the pristine images. Three levels of blur, three levels of noise, and combinations of blur and noise produced a total of 252 distorted images.

The subject test procedure was written using Matlab and the PsychToolbox [49], [50]. Each subject was first presented with a training session in which 10 images were shown before the main testing session, to give them an understanding of how to score images. Two testing sessions were performed with each session containing a unique set of 126 images. Subjects were presented with a single stimulus image for 10 seconds as depicted in Fig. 12. At the end of the 10 seconds, a continuous sliding quality bar with the labels "Bad", "Poor", "Fair", "Good", or "Excellent" was presented, as shown in Fig. 13.

Each image was scored by 24 subjects with each score discretized to integers on [0, 100]. In order to account for differences in image content, we computed the Difference

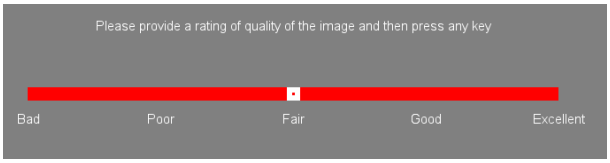


Fig. 13. Sliding Quality Bar.

Mean Opinion Scores (DMOS). Let s_{ijk} be the opinion score given by subject i, on image j during session $k = \{1, 2\}$. Then the difference score for subject i, image j, and session k is given by

$$d_{ijk} = s_{ij_{ref}k} - s_{ijk}, \quad s_{ij_{ref}k} \neq s_{ijk},$$

where $s_{ij_{ref}k}$ is the score given to the (hidden) pristine image corresponding to the distorted one. The difference scores from each session were then converted to Z-scores:

$$z_{ijk} = \frac{d_{ijk} - \mu_{ik}}{\sigma_{ik}}$$

where

$$\mu_{ik} = \frac{1}{N_{ik}} \sum_{j=1}^{N_{ik}} d_{ijk}$$

and

$$\sigma_{ik} = \sqrt{\frac{1}{N_{ik} - 1} \sum_{j=1}^{N_{ik}} (d_{ijk} - \mu_{ik})^2}$$

and where N_{ik} is the number of test images seen by subject i in session k.

The subject rejection procedure specified in the ITU-R BT 500.11 recommendation is useful for discarding scores from unreliable subjects. Z-scores are considered normally distributed if their kurtosis falls between the values of 2 and 4. The recommendation is to reject if more than 5 percent of the Z-scores lie outside two standard deviations of the mean. Using this procedure, all except one subject was found to be acceptable. The one outlier chose the same value of 50 for all images. Thus only one subject was rejected [50], [51].

After the subject rejection procedure, the values of z_{ijk} fell into the range $[-3, 3]$. A linear rescaling was used to remap the scores onto $[0, 100]$ using

$$z'_{ij} = \frac{100(z_{ij} + 3)}{6}$$

Finally the Difference Mean Opinion Score (DMOS) of each image was computed as the mean of the $M = 24$ rescaled Z-scores:

$$\text{DMOS}_j = \frac{1}{M} \sum_{i=1}^{M} z'_{ij}.$$

A plot of the histogram of the DMOS scores is shown in Fig. 14, indicating a reasonably broad distribution of the DMOS scores.

Table VI shows the Spearman's Rank Correlation Coefficient (SRCC) and (Pearson's) linear correlation coefficient (LCC) between subjective scores and model predictions for NR feature groups. The results were computed using 1000 iterations of randomly sampled training and testing groups. As in the previous experiments, 80 percent of the data was used for training and the remainder for testing. Care was taken to not overlap training and testing on the same content in any iteration since such an overlap could inflate performance results by training on the content rather than distortion. An SVR was used to fit the NSS feature parameters to the DMOS scores.

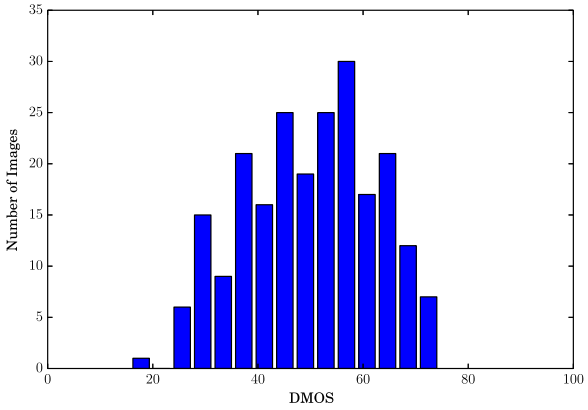


Fig. 14. Histogram of DMOS scores.

TABLE VI
MEDIAN SRCC AND LCC BETWEEN DMOS AND PREDICTED DMOS MEASURED OVER 1000 ITERATIONS

NR Feature Set	Median SRCC	Median LCC
$f + pp + pd + sp$	0.815	0.820
$f + pp + pd$	0.794	0.809
$f + pp$	0.809	0.817
$f + pd$	0.727	0.742
f	0.714	0.736
pp	0.794	0.809
pd	0.696	0.732
sp	0.825	0.828
IQIs	0.726	0.705
Ro, L_1	0.135	0.189
Ro, L_2	0.162	0.221
ERo, L_1	0.570	0.576
ERo, L_2	0.616	0.667

We observe that the sp group of features provided the highest correlation with the human subjective scores, being only a slight improvement over the BRISQUE model, $f + pp$. Combining feature groups with sp produces worse correlation indicating possible overfitting of the training set. For these blur and AWN distortions, the directional feature groups provided the highest correlation with DMOS scores. The IQIs and NU distortion-based models provided comparatively low correlation. The proposed perceptual models provide excellent prediction of human opinion scores.

F. Local Distortion Maps

Image maps that highlight locally distorted regions can be useful not only for detecting the presence and location of particular distortions such as hotspots or halos, but also for detecting any unknown (combination of) distortions. It is

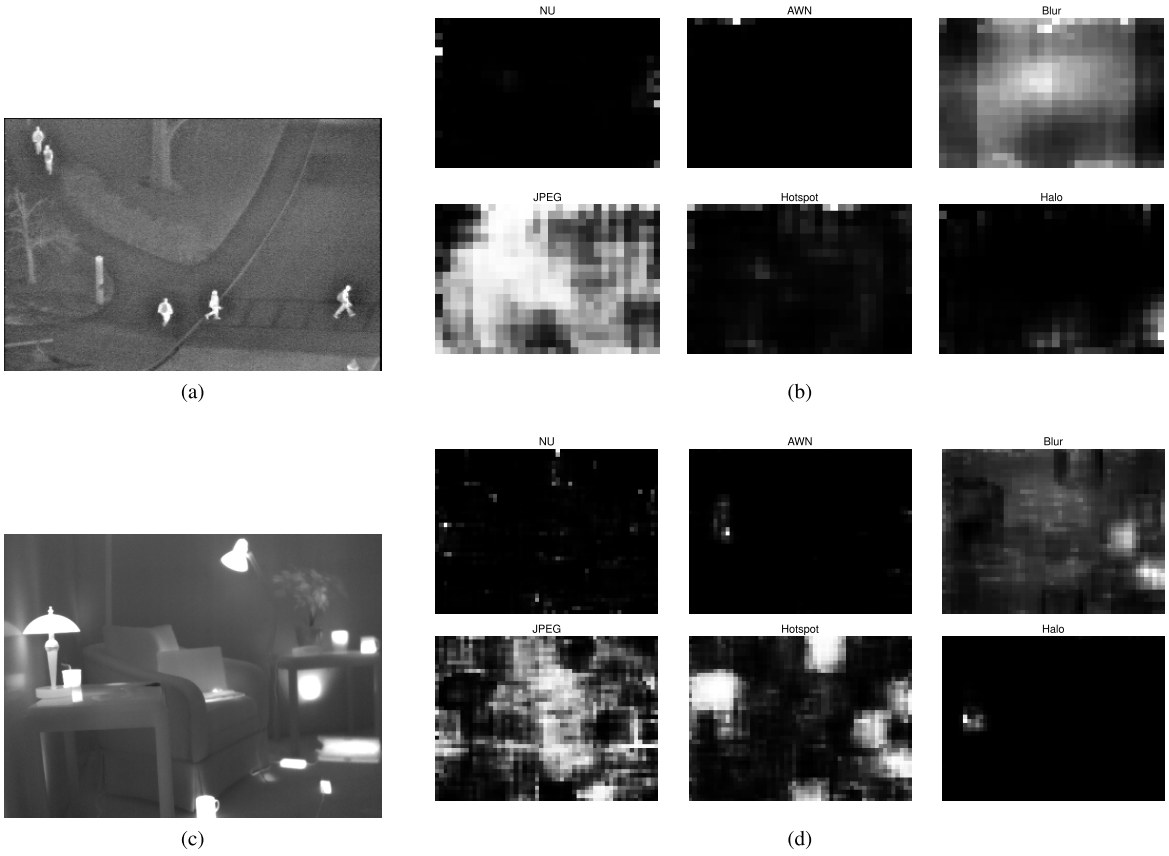


Fig. 15. Heatmaps represent probabilities computed at a particular spatial location using an SVM classifier trained on two classes. Class 1 represents natural IR images and Class 2 represents one of NU, AWN, Blur, JPEG, Hotspot, or Halo. Six SVM classifiers are trained to produce these six heatmaps. Brighter regions represent higher probabilities. Each heatmap has been contrast stretched for visibility. (a) Source Image with "Halo effect". (b) Heatmaps. (c) Source Image with hotspots. (d) Heatmaps.

possible to automatically find local distorted regions of LWIR images using NSS-based features.

A distortion map can be generated using a sliding window to capture ROIs from the image being analyzed. We used a 96×96 sliding square window scanned along the image in 12 pixel step increments (strides). Thus each captured ROI overlapped with 87.5 percent of the last ROI in sequence. Each ROI was classified using multiple probabilistic SVCs, one per distortion type, to determine the probability that the ROI was distorted by a particular distortion. The probabilities of distortion were gathered and mapped into an image that highlights distorted areas. Example distortion maps are shown in Fig. 15. Some distortion maps, such as on JPEG distorted images appear in Fig. 15 to provide false positives, but this is an artifact of displaying full-contrast stretched low probability values. For detecting unknown distortions, one could use a localized variant of the method in VIID (as described in Sec. III-A), since the NSS of natural IR images are highly predictable.

IV. Conclusion and Future Work

LWIR images possess statistical regularities similar to those of visible light images. Furthermore, the NSS of LWIR images are distinct from the NSS of visible light images, as shown through both statistical analysis and the aforementioned VIID classifier. These powerful NSS descriptors provide an accurate global distortion level estimate, as demonstrated in the cases of non-uniformity and white noise. Again, these descriptors are perceptually driven, and thus correlate well with human task performance and human subjective test results. Lastly, they can be used to build a local distortion detector, which can aid further work in designing image enhancement and compression methods. Each of these described tasks is improved by the addition of NSS features.

Extensions to this work may include study of geometric distortions, infrared reflections, and radiometric distortions. Such a study may prove to be fruitful for industry given the manufacturing limitations associated with producing LWIR imagers.

The NSS of LWIR videos are also of great interest. The LWIR videos used in surveillance could be modeled and studied to improve compression techniques, object detection algorithms, and overall video quality. The common visible light video compression formats including H.264/MPEG4 could be studied in light of thermal sensitivity requirements. Thermal variance, which is separate from and poorly represented by models of motion, would be worthy of analysis as well.

Acknowledgment

The authors thank NIST for providing their pristine LWIR images. They also acknowledge Jennifer Kaser at the

University of Texas at Austin, Dr. Nigel of the University of Toronto, and Dr. James and Dr. Vinay of Ohio State University for providing LWIR images for study.

REFERENCES

[1] C. Ibarra-Castanedo *et al.*, "Comparative study of active thermography techniques for the nondestructive evaluation of honeycomb structures," *Res. Nondestruct. Eval.*, vol. 20, no. 1, pp. 1–31, 2009.

[2] Aeryon Labs Inc. (2014). [Online]. Available: http://www.aeryon.com/

[3] F. K. Amon, N. P. Bryner, and A. Hamins, *Thermal Imaging Research Needs for First Responders: Workshop Proceedings*. Washington, DC, USA: U.S. Government Printing Office, 2005.

[4] F. Amon and A. Lock, "Evaluation of image quality of thermal imagers used by the fire service," U.S. Dept. Commerce, Nat. Inst. Standards Technol., Washington, DC, USA, NIST Tech. Note 1630, 2009.

[5] F. K. Amon, D. Leber, and N. Paulter, "Objective evaluation of imager performance," in *Proc. 5th Int. Conf. Sens. Technol.*, 2011, pp. 47–52.

[6] N. A. Diakides and J. D. Bronzino, *Medical Infrared Imaging*. Boca Raton, FL, USA: CRC Press, 2007.

[7] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer. A*, vol. 4, no. 12, pp. 2379–2394, 1987.

[8] D. J. Tolhurst, Y. Tadmor, and T. Chao, "Amplitude spectra of natural images," *Ophthalmic Physiol. Opt.*, vol. 12, no. 2, pp. 229–232, 1992.

[9] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[10] D. L. Ruderman, "The statistics of natural images," *Netw., Comput. Neural Syst.*, vol. 5, no. 4, pp. 517–548, 1994.

[11] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annu. Rev. Neurosci.*, vol. 24, no. 1, pp. 1193–1216, 2001.

[12] B. A. Olshausen and D. J. Field, "Natural image statistics and efficient coding," *Netw., Comput. Neural Syst.*, vol. 7, no. 2, pp. 333–339, 1996.

[13] J. M. Mooney, F. D. Sheppard, W. S. Ewing, J. E. Ewing, and J. Silverman, "Responsivity nonuniformity limited performance of infrared staring cameras," *Opt. Eng.*, vol. 28, no. 11, p. 281151, 1989.

[14] J. M. Mooney, "Effect of spatial noise on the minimum resolvable temperature of a staring sensor," *Appl. Opt.*, vol. 30, no. 23, pp. 3324–3332, 1991.

[15] J. M. López-Alonso, J. Alda, and E. Bernabéu, "Principal-component characterization of noise for infrared images," *Appl. Opt.*, vol. 41, no. 2, pp. 320–331, 2002.

[16] J. E. Pezoa and O. J. Medina, "Spectral model for fixed-pattern-noise in infrared focal-plane arrays," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. Berlin, Germany: Springer-Verlag, 2011, pp. 55–63.

[17] F. Pérez, M. Nova, J. E. Pezoa, M. Figueroa, and S. N. Torres, "Spatial and frequency domain metrics for assessing fixed-pattern noise in infrared images," in *Proc. SBMO/IEEE MTT-S Int. Microw. Optoelectron. Conf.*, Aug. 2013, pp. 1–5.

[18] F. Pérez, J. E. Pezoa, M. Figueroa, and S. N. Torres, "Empirical frequency domain model for fixed-pattern noise in infrared focal plane arrays," *Infr. Phys. Technol.*, vol. 67, pp. 413–426, Nov. 2014.

[19] J. Y. Kaser. (2013). *Utilizing Natural Scene Statistics and Blind Image Quality Analysis of Infrared Imagery*. [Online]. Available: http://repositories.lib.utexas.edu/handle/2152/22601, accessed Dec. 12, 2013.

[20] T. Goodall and A. C. Bovik, "No-reference task performance prediction on distorted LWIR images," in *Proc. IEEE Southwest Symp. Image Anal. Interpretation*, Apr. 2014, pp. 89–92.

[21] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[22] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[23] M. M. Hayat, S. N. Torres, E. Armstrong, S. C. Cain, and B. Yasuda, "Statistical algorithm for nonuniformity correction in focal-plane arrays," *Appl. Opt.*, vol. 38, no. 5, pp. 772–780, 1999.

[24] C. S. Martina, S. N. Torresa, and J. E. Pezoa, "An effective reference-free performance metric for non-uniformity correction algorithms in infrared imaging system," in *Proc. 20th Annu. Meeting Laser Electro-Opt. Soc.*, Lake Buena Vista, FL, USA, 2007, pp. 21–25.

[25] J. V. Anda. (2008). "Uncooled detectors for thermal imaging cameras," FLIR, Tech. Note 16. [Online]. Available: http://www.flir.com/uploadedFiles/Eurasia/Cores_and_Components/Technical_Notes/uncooled%20detectors%20BST.pdf

[26] C. M. Hanson, H. Beratan, R. A. Owen, M. Corbin, and S. McKenney, "Uncooled thermal imaging at texas instruments," *Proc. SPIE*, vol. 1735, pp. 17–26, Dec. 1992.

[27] G. C. Holst, *CCD Arrays, Cameras, and Displays*. Winter Park, FL, USA: SPIE, 1998.

[28] J. W. Davis and V. Sharma, "Robust background-subtraction for person detection in thermal imagery," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshop*, Jun. 2004, p. 128.

[29] I. B. Schwartz, K. A. Snail, and J. R. Schott, "Infrared halo effects around ships," DTIC Document, NRL Memorandum Rep. 5529, 1985.

[30] J. L. Rowe, *The Impact of Thermal Imaging Camera Display Quality on Fire Fighter Task Performance*. Ann Arbor, MI, USA: ProQuest, 2008.

[31] N. J. W. Morris, S. Avidan, W. Matusik, and H. Pfister, "Statistics of infrared images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–7.

[32] J. W. Davis and M. A. Keck, "A two-stage template approach to person detection in thermal imagery," in *Proc. 7th IEEE Workshops Appl. Comput. Vis.*, Jan. 2005, pp. 364–369.

[33] J. W. Davis and V. Sharma, "Background-subtraction using contour-based fusion of thermal and visible imagery," *Comput. Vis. Image Understand.*, vol. 106, no. 2, pp. 162–182, 2007.

[34] *Infrared Image Quality Assessment at Live*. [Online]. Available: http://live.ece.utexas.edu/research/IR/index.htm, accessed Nov. 20, 2014.

[35] Y. Zhang and D. M. Chandler, "An algorithm for no-reference image quality assessment based on log-derivative statistics of natural scenes," *Proc. SPIE*, vol. 8653, p. 86530J, Feb. 2013.

[36] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 9, pp. 891–906, Sep. 1991.

[37] E. P. Simoncelli and D. J. Heeger, "A model of neuronal responses in visual area MT," *Vis. Res.*, vol. 38, no. 5, pp. 743–761, 1998.

[38] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[39] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 1, pp. 52–56, Feb. 1995.

[40] N.-E. Lasmar, Y. Stitou, and Y. Berthoumieu, "Multiscale skewed heavy tailed model for texture analysis," in *Proc. IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 2281–2284.

[41] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. *LIVE Image Quality Assessment Database Release 2*. [Online]. Available: http://live.ece.utexas.edu/research/quality, accessed May 2014.

[42] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[43] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[44] *The DRIQ Database*. (2012). [Online]. Available: http://vision.okstate.edu/index.php?loc=driq, accessed Jul. 1, 2015.

[45] C. Vu, T. Phan, and D. Chandler, "Can current image quality assessment algorithms predict visual quality of enhanced images?" to be published.

[46] A. Zarić *et al.*, "VCL@FER image quality assessment database," *AUTOMATIKA*, vol. 53, no. 4, pp. 344–354, 2012.

[47] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[48] R. De Maesschalck, D. Jouan-Rimbaud, and D. L. Massart, "The Mahalanobis distance," *Chemometrics Intell. Lab. Syst.*, vol. 50, no. 1, pp. 1–18, 2000.

[49] D. H. Brainard, "The psychophysics toolbox," *Spatial Vis.*, vol. 10, no. 4, pp. 433–436, 1997.

[50] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1427–1441, Jun. 2010.

[51] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document ITU-R, Rec. BT.500-11, Int. Telecommun. Union, Geneva, Switzerland, 2002.

Todd Richard Goodall received the B.S. degree in computer engineering and the M.S.E. degree in electrical and computer engineering from Clemson University, Clemson, SC, USA, in 2012 and 2014, respectively. He is currently pursuing the Ph.D. degree in electrical and computer engineering with the Laboratory for Image and Video Engineering, The University of Texas at Austin.

He was a recipient of the NDIA UWD Academic Fellowship from 2012 to 2013. His research interests include the statistical modeling of images and videos, design of image and video quality assessment algorithms, visual perception, and computational vision.

Alan Conrad Bovik (F'96) currently holds the Cockrell Family Endowed Regents Chair in Engineering at The University of Texas at Austin, where he is Director of the Laboratory for Image and Video Engineering. He is a faculty member with the Department of Electrical and Computer Engineering and the Institute for Neuroscience. His research interests include image and video processing, computational vision, and visual perception. He has authored over 750 technical articles in these areas and holds several U.S. patents. His publications have been cited more than 45 000 times in the literature, his current H-index of 75, and he is listed as a Highly-Cited Researcher by Thompson Reuters. His several books include the companion volumes *The Essential Guides to Image and Video Processing* (Academic Press, 2009).

Dr. Bovik received a Primetime Emmy Award for Outstanding Achievement in Engineering Development from the Academy of Television Arts and Sciences (The Television Academy) in October 2015, for his contributions to the development of video quality prediction models which have become standard tools in broadcast and post-production houses throughout the television industry. He has also received a number of major awards from the IEEE Signal Processing Society, including: the Society Award (2013); the Technical Achievement Award (2005); the Best Paper Award (2009); the Signal Processing Magazine Best Paper Award (2013); the Education Award (2007); the Meritorious Service Award (1998); and (co-author) the Young Author Best Paper Award (2013). He was also named recipient of the Honorary Member Award of the Society for Imaging Science and Technology for 2013, received the SPIE Technology Achievement Award for 2012, and was the IS&T/SPIE Imaging Scientist of the Year for 2011. He is also a recipient of the Hocott Award for Distinguished Engineering Research (2008) and the Joe J. King Professional Engineering Achievement Award (2015) from the Cockrell School of Engineering at The University of Texas at Austin (2008), and the Distinguished Alumni Award from the University of Illinois at Champaign-Urbana (2008). He co-founded and was the longest-serving Editor-in-Chief of the IEEE Transactions on Image Processing (1996-2002), created and served as the General Chairman of the First IEEE International Conference on Image Processing, held in Austin, Texas, in 1994, along with numerous other professional society activities, including the Board of Governors, IEEE Signal Processing Society from 1996 to 1998; the Editorial Board of the IEEE Proceedings, 1998-2004; and a Series Editor of Image, Video, and Multimedia Processing (2003-present, Morgan and Claypool Publishing Company).

Dr. Bovik is a registered Professional Engineer in the state of Texas and is a frequent consultant to legal, industrial, and academic institutions.

Nicholas G. Paulter, Jr. (F'10) was the Leader of the High-Speed Pulse Metrology Project at National Institute of Standards and Technology (NIST), Gaithersburg, MD, and a Program Manager at OLES from 1990 to 2005. He was a Program Manager of the Detection, Enforcement, and Inspection Program of the Law Enforcement Standards Office and is the Group Leader of the Security Technologies Group at NIST. He develops and oversees metrology programs related to concealed-weapon detection, through-wall surveillance, traffic control devices, imaging for security and emergency response applications, biometrics for identification, and weapon output characterization. In this role, he has initiated two NIST programs in submillimeter-wave concealed weapon imaging, facilitated the development of visible-light hyper-spectral image projection, initiated an NIST-centric imaging metrology program for security and emergency applications, started two through-wall surveillance/sensing metrology activities, and started a program for the characterization of the high-voltage output of electroshock weapons. He received an NIST Bronze Medal in 2003 for his work in developing minimum performance requirements for metal detectors. In that capacity, he developed several high-speed electrical pulse generation and sampling systems, electrooptic-based measurement systems, and short optical pulse laser systems for use in pulse metrology. Moreover, during his tenure as the Project Leader, the metrology services provided by his project became the best in the world.

Mr. Paulter received the B.S. degree in biological sciences from the University of California, Santa Barbara, in 1980, and the M.S. degrees in chemistry and electronics engineering from the University of New Mexico and the University of Colorado at Boulder, in 1988 and 1990, respectively. He was with the Los Alamos National Laboratory, Los Alamos, NM, from 1980 to 1989, and was involved in the study of fast electrical and optical phenomena. He has authored or co-authored over 100 peer-reviewed technical articles and provided numerous presentations at a variety of technical conferences. He is a Commerce Science and Technology Fellow from 2008 to 2009.